Write a summary of the exploratory data analysis above. What numerical or categorical features were in the data? Was there any pattern suggested of a relationship between state and ticket price? What did this lead us to decide regarding which features to use in subsequent modeling? What aspects of the data (e.g. relationships between features) should you remain wary of when you come to perform feature selection for modeling? Two key points that must be addressed are the choice of target feature for your modelling and how, if at all, you're going to handle the states labels in the data.

The exploratory data analysis was used for two main tasks: comparison of state specific features against average price and comparison generally of features vs. average price.

To compare states against each other it was necessary to compute relevant ratios of resorts features such as skiable area and number of resorts against state population and size. These ensures that the resorts are the driving force behind the comparison and not the qualities of the state superimposed over them. We still need to be diligent to not forget peculiarities of states such as New York that has a considerable amount of wealthy persons within reasonable weekend driving distance to many of the resorts as opposed to Montana where most visitors would need to fly in or drive excessive distances. Generally, it was found that there was no consistent relationship between state and average weekend ticket price. We should keep the data labels for now though we may drop them for some analyses that require sub setting. The state names are a good way to represent the data at the end in a presentation as they are more recognizable to most people than the 330 individual resort names.

The general comparison of resort features provided more promising results for modeling pricing. Thru Seaborn heatmapping and scatterplots we were able to determine the following features had relatively high correlation to weekend ticket price: vertical drop, fast quads, total chairs, total runs and snowmaking acres. The relationships still require further exploration to determine their limits. For instance, it appears that the addition of the first fast quad has considerably more upside than the addition of subsequent fast quads. Also, while summit and base elevation seem to have a slight correlation to weekend price, the vertical drop feature which combines them shows a string correlation. We will move forward with these features as the targets to study further for their relationship to weekend ticket price.